EROS: A Protein Docking Algorithm Using a Quaternion Pi-Ball Representation for Exhaustive and Accelerated Exploration of 3D Rotational Space

Maria Elisa Ruiz Echartea, Isaure Chauvot de Beauchêne, David Ritchie

EROS: A Protein Docking Algorithm Using a Quaternion π-Ball Representation for Exhaustive and Accelerated Exploration of 3D Rotational Space

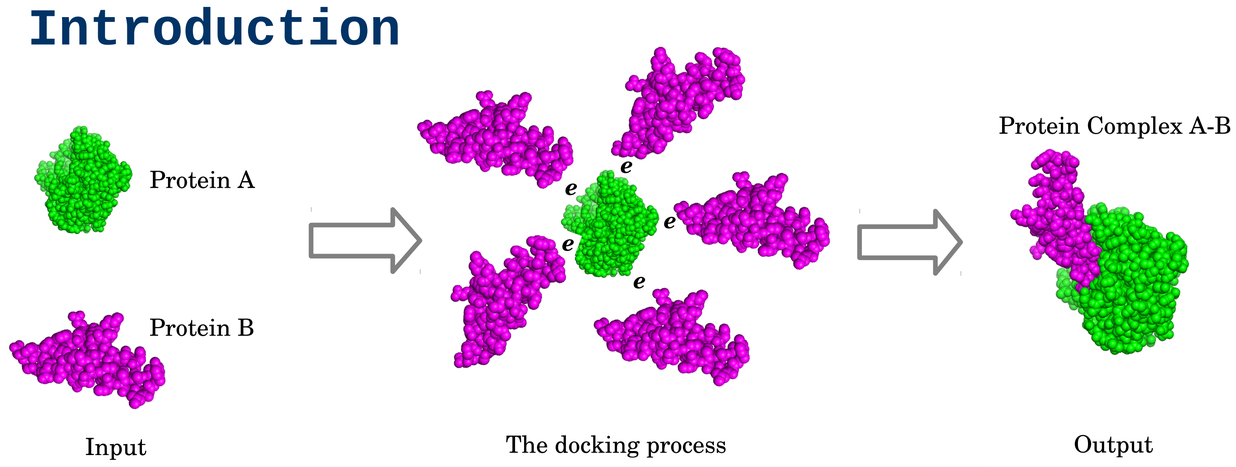
Maria Elisa Ruiz-Echartea[1], Isaure Chauvot de Beauchêne[1], David W Ritchie [1]

1. Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

1. Introduction

Proteins are involved in many essential cellular processes of living organisms. Proteins form macro complexes joining themselves to other proteins to carry out these processes. Therefore, to know the 3D structures of such complexes is of biomedical interest.

Protein-protein docking algorithms aim to predict how two proteins interact with each other to form a 3D complex. Docking algorithms need to fulfill two main tasks: (1) sampling all the possible relative positions of the two proteins and (2) computing the interaction energy at each position to find the minimum energy (= the best solution). Obtaining the interaction energy is a computationally expensive task. We are developing a new algorithm based on the ATTRACT coarse-grained force-field [1] and using a quaternion π-ball representation to accelerate the search of the 3D rotational space.



2.The Quaternion π-Ball

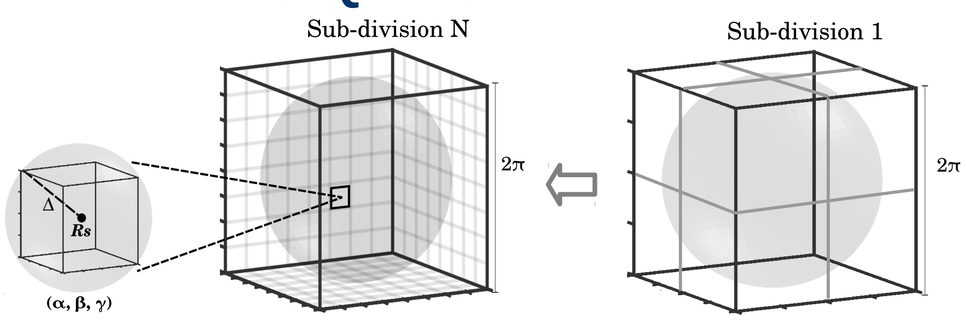


Fig 1. The π-ball representation.

The 3D rotational space is represented by a π-ball contained inside a cube. Any point within the π-ball may be mapped to an Euler rotation defined by the three Euler rotation angles, (α, β, γ). Points within the π-ball may be represented as a unit quaternion, Q = (cos(θ/2), sin(θ/2)u), where u is a unit vector from the origin.

A series of sample rotations is generated by dividing the cube containing the π-ball into 8 sub-cubes, and recursively sub-dividing each such cube into 8 smaller cubes until a given threshold cube dimension is reached. Each cell contains a subspace of similar rotations of the rotations space, since they are near each other.

3. Divide and Conquer

The set of all possible sample rotations from the π-ball cube centres ***Rc*** and their radii **Δ** may be collected as a set of nodes in a 3D "search tree" data structure.

Each node can be tested to know if such subspace leads to **forbidden orientations** (i.e. leading to atomic clashes). If yes, the node leads to a forbidden orientation so it is labeled along with its descendants as forbidden.

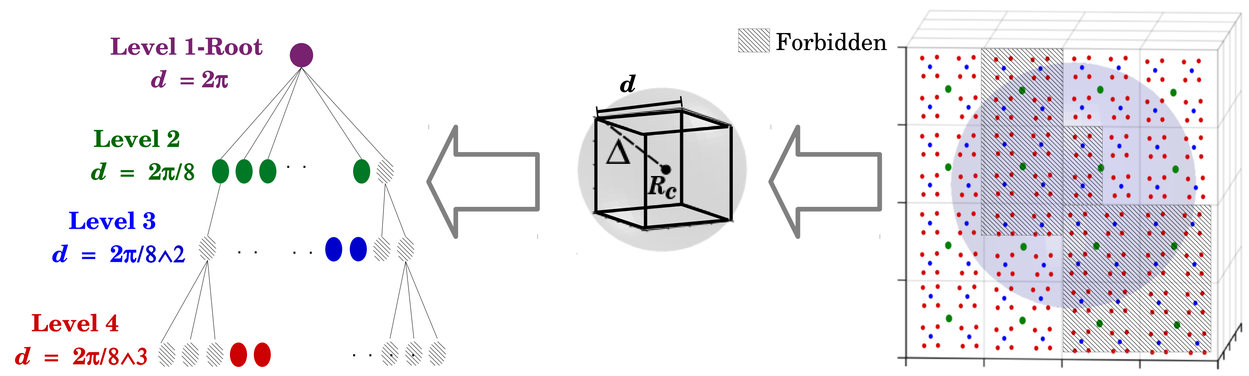


Fig 2. a) Tree structure representing the cells of the π-ball and the π-ball with some shadow areas representing the forbidden rotations.

4. Application to Protein Docking

Based on a preliminary study, we found that a large number of protein interfaces contain at least one pair of atoms at almost their optimal distance according to an empirical force field. Therefore, our idea is to perform a series of restricted docking searches in which a pair of surface beads from each protein is placed at the coordinate origin (see Fig. 4).

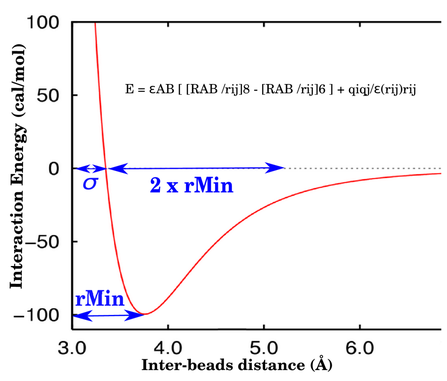


Fig 3. ATTRACT force field describing the interaction energy for a given pair of beads.

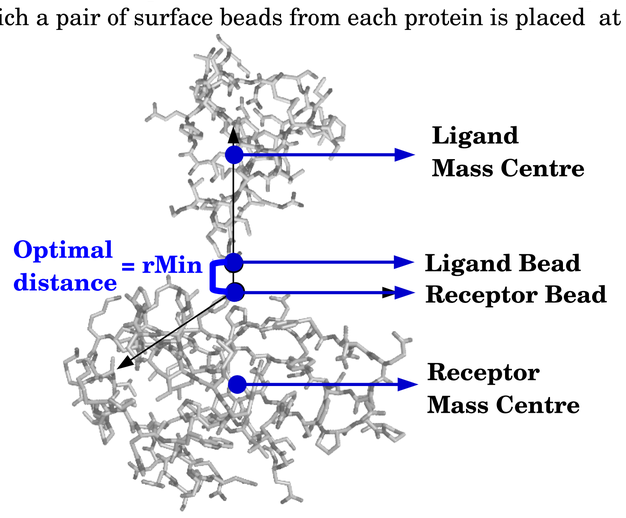


Fig 4. Example of an initial starting position for a 3D rotational search.

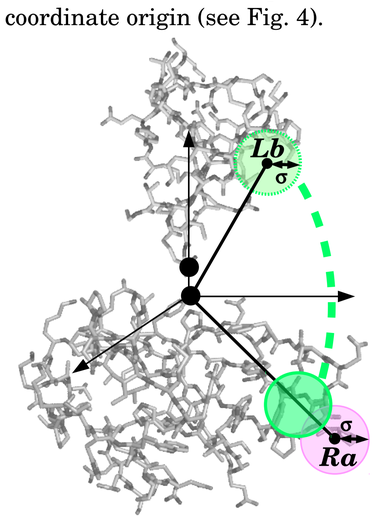


Fig 5. Two beads aligned to find if they are clashing.

In order to detect the bead pairs that can lead to steric clashes upon ligand rotation, we compute the differences between the distances. Thus, for all (a, b) in (Receptor, Ligand):

If $|Ra - Lb| > \sigma ab$, where σab is the clashing distance for atom types a and b, see Fig. 3 and 5,
then beads a and b will never clash
else
we have to examine in more detail if the beads might overlap or not (see Fig. 5 and 6).

Then, the tree is walked to label as forbidden those rotations of the search space that intersect some orientation of the list. If a bead comes close enough to receptor bead *a* to always give a clash, then that rotation and all the rotations in the cube's descendants are labeled as forbidden (see Fig. 6).

In the last stage, the tree is walked to compute energies in order to find the best solutions. The π-ball is used as a rotational map. Therefore, many trial rotations are discarded and the rotational search space is pruned to achieve an improvement in efficiency.

In our experiments we use a minimum cell radius of 7.5 degrees, which divides the π-ball into 299,593 cells. We observe that about 96% of these are forbidden, leaving only 4% of the search space where interaction energies have to be computed.

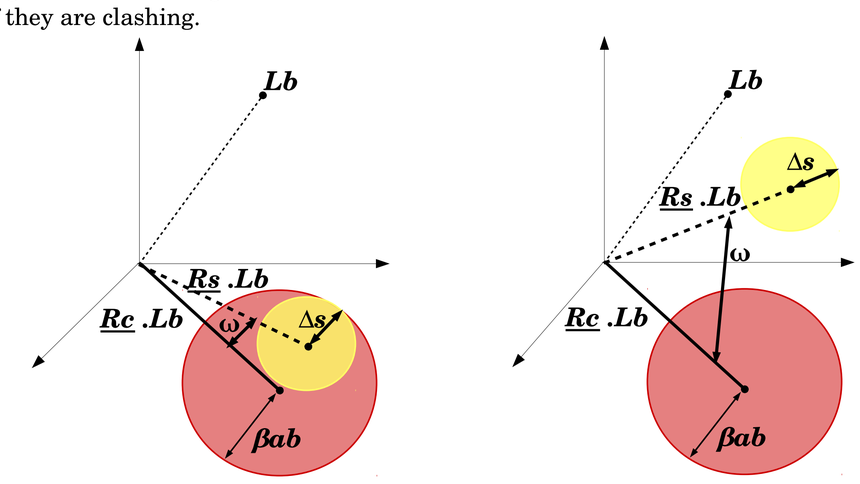
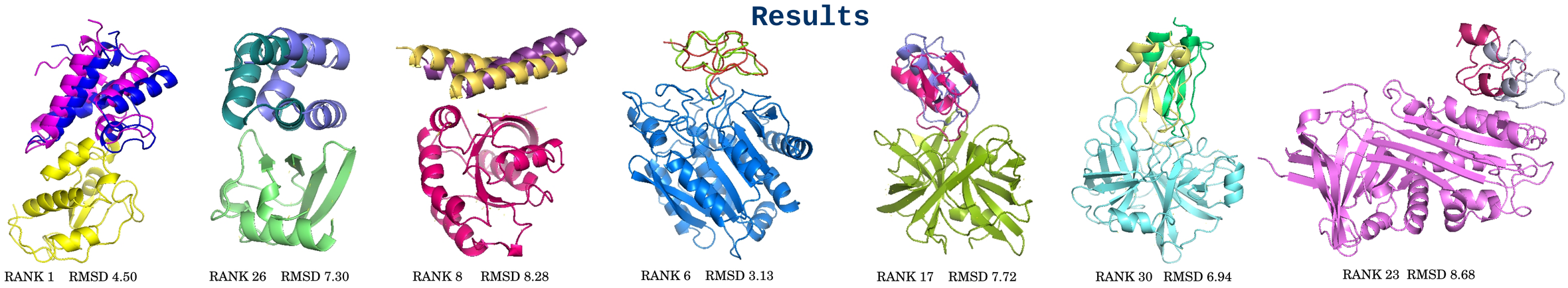


Fig 6. Possible cases when a rotation moves a bead.

Case 1) ω < β - Δs => Every rotation within cube s cause a steric clash.

Case 2) ω > β + Δs => No rotation within cube s can cause a steric clash.

Results



RANK 1 RMSD 4.50 | RANK 26 RMSD 7.30 | RANK 8 RMSD 8.28 | RANK 6 RMSD 3.13 | RANK 17 RMSD 7.72 | RANK 30 RMSD 6.94 | RANK 23 RMSD 8.68

The Root-Mean-Square Deviation (RMSD) is a measure of the average distance between the atoms of two superimposed proteins. In practice, to evaluate the efficiency of a docking algorithm, the RMSD is used to evaluate the difference between a calculated and an experimentally determined 3D structure. Above, we show the rank and RMSD of some of the solutions found by our algorithm.

Conclusions

The rotational space can be labeled in order to identify unfavorable rotations before computing any energy. Thus, the labeled π-ball can be used as a rotational map avoiding to compute energies at useless rotations. Therefore, the exploration of the rotational space is done is an efficient way. The exploration is exhaustive since the complete rotational space is analyzed and labeled.

Acknowledgements

Maria Elisa Ruiz Echartea is supported by a doctoral bursary (CORDI-S) from Inria Nancy Grand Est.

References

1. Zacharias, M. Protein Science, 12, 1271-1282 (2003).